MEASURING THE IMPACT OF THE QUALITY OF MATHEMATICS INSTRUCTION ON STUDENT ACHIEVEMENT DURING THE MIDDLE SCHOOL YEARS: PILOT STUDY

APRIL 2004

Pamela B. Stazesky, Ph.D.
Associate Director

Delaware Education
R&D
Research & Development
CENTER

Delaware Education Research & Development Center
University of Delaware
Newark, DE 19716

T04.005.2

INTRODUCTION

All projects are based on theories, although often unstated, of how and why they should "work" (Weiss, 1995). Theory-based evaluation provides a useful framework for formalizing the logic of the theories underlying a project and in guiding the determination of measurement points during the evaluation (Aronson, Mutchler, & Pan, 1998). Examining theories on which a project is based aids in determining what data should be collected as well as when during the project lifecycle the data should be collected. The ultimate goal of the Delaware Exemplary Mathematics 6-12 Curriculum Implementation (DEMCI) project funded by the National Science Foundation [Grant No. 9819592] is to increase students' mathematics achievement. Theory-based evaluation methods were used to document why the project staff believes this intervention will result in an increase in learning and to specify what data must be collected to determine if intervention results support these theories. The critical theory behind this project is that through providing standards-based instructional materials and professional development to teachers, the quality of mathematics instruction will improve and, consequently, student learning will also improve. With this theory in mind, data elements were identified that would aid in determining if this theory was acceptable. Appendix A shows a simplified theory-based logic map for the DEMCI project. The fundamental research question this study will attempt to answer is to what extent does the quality of standards-based mathematics teaching in the middle grades influence student mathematics achievement.

DEMCI Project

Beginning in 1999, all middle and high school mathematics teachers (grades 5-12) in 16 of the 19 school districts in Delaware were invited and strongly encouraged to receive comprehensive training in the curriculum they have selected – Mathematics in Context, Connected Mathematics, Contemporary Mathematics in Context, or MATH Connections. The belief if only the most receptive teachers at a school are trained in a curricular innovation, then the initiative will likely become marginalized and disappear in the relative short-term has been adopted by this project. In addition, this project believes that if teachers are required to adopt a new curriculum without sufficient training, they and their students are at great risk of a failed implementation. Furthermore, teachers with a weak mathematics background will also jeopardize the curriculum implementation. Since the inception of the grant, one district has chosen to withdraw and another school district has recently taken a hands-off approach permitting each individual school to renegotiate their involvement. In another district, the middle schools have chosen to withdraw; however, the high schools have remained with the project.

While the much of the professional development for this project occurred during the intensive week-long summer institutes, teachers may also participate in after-school study groups, one-on-one consultations with on-site math specialists as well as periodic special professional development sessions on various topics such as working with special education students in a standards-based mathematics classroom. It is the goal that each teacher in the project will attend three summer institutes including a culminating institute during the summer of 2002. These summer institutes will provide 150 hours of professional development for each teacher. The collaborative work between the teachers and the math specialists on-site are designed to add at least another 40 hours of professional development each year.

Central to the infrastructure of the project is a team of math specialists who work as a team to provide instructional support and mentoring to teachers in participating schools. This team is drawn from the mathematics faculties of the participating school districts as teachers on special assignment for two or three years. These specialists spend most of their time working with teachers in participating schools, but also meet as a team with the principal investigators on a regular basis to plan, problem solve, and debrief.

METHODS

This research study relied on two data sources: a) results from the mathematics portion of the state assessment, the Delaware Student Testing Program (DSTP), which is administered every spring to students in grades 3, 5, 8, and 10 throughout the state, and b) data regarding the quality of mathematics instruction students received.

To determine if the quality of mathematics instruction students received in 8th grade influences their score on the mathematics portion of the DSTP, we compared the performance of 8th grade students on the mathematics portion of the DSTP for three different groups of students – those who received high quality mathematics instruction, those who received moderate quality mathematics instruction, and those who received poor quality mathematics instruction in 8th grade. However, because research indicates that previous achievement has a strong influence on future achievement (Keith & Lichtman, 1992; Vollmer, 1986; Young, Reynolds, & Walberg, 1996), students' performance on the mathematics portion of the 5th grade DSTP was also included in this study. Therefore, we conducted a two-way analysis of variance (ANOVA) using the performance of 8th grade students on the mathematics portion of the DSTP as the dependent variable. We used each student's performance level (well below, below, meets, exceeds, and distinguished) from the 5th grade mathematics DSTP and the quality of instruction (high, moderate, or poor) he/she received in 8th grade mathematics as the independent variables. This analysis resulted in 15 different groups of students.

Sample

During the 2002-03 school year, a sample of 39 eighth grade teachers in Delaware was selected for participation in this study. These teachers were selected because of their school's participation in the NSF-funded Local Systemic Change Initiative, known in Delaware as the Delaware Exemplary 6-12 Mathematics Curriculum Implementation (DEMCI) project. Only middle schools that had participated in this grant were included in this study, i.e., middle schools that had adopted either Mathematics in Context or Connected Mathematics curriculum materials. Therefore, middle schools classrooms in which a traditional Algebra I text in 8th grade was used were not included in this study. In addition, teachers in these schools were included only if a Secondary Mathematics Specialist (SMS) working with the DEMCI project had specific knowledge about the quality of their teaching.

Because in a typical school students are not randomly assigned to classrooms, all 8th grade students (n = 2,102) rather than a random sample of students from each of these 39 teachers were selected for inclusion in this study. While most of the students in this study (87%) were in fifth grade during the 1999-2000 school year, this study also included students who had been retained one time between 5th and 8th grade. In addition, only students for whom complete data were available were included. That is, all of the 8th grade students in these classrooms for whom a valid 5th grade and a valid 8th grade DSTP score were available were included in this study. However, one group of students, English Language Learners, was excluded from this analysis. Students in the sample represent children from seven school districts and fourteen middle schools across all three counties in Delaware. The demographics of the student sample used for this study as compared to 8th grade students throughout the state of Delaware are listed in Table 1. The percentages listed in parentheses represent the population of 8th grade students in Delaware public schools during the 2002-03 school year. The percentages not listed in parentheses represent the sample of 8th grade students in Delaware public schools during the 2002-03 school year who participated in this study. These data indicate that the students in the sample are very similar to the students across the state in all categories except special education. The sample in this study has slightly fewer students receiving special education services (9%) than the state as a whole (15%).

Table 1

Demographic Characteristics of Delaware 8th Grade Students

Gender			Male 50% (52%)		Female 50% (48%)
Special Education Services			Yes 9% (15%)		No 91% (85%)
Title I Services			Yes 17% (17%)		No 82% (83%)
	Caucasian	African American	Hispanic	Asian or Pacific Islander	American Indian or Alaskan
Race	63% (60%)	30% (32%)	5% (6%)	1% (2%)	<1% (<1%)

Delaware Student Testing Program – Mathematics

The Delaware Student Testing Program (DSTP) is a written assessment administered every spring to students attending public schools in grades 3, 5, 8, and 10 throughout the state of

Delaware. This assessment was implemented for the first time in the spring of 1998 and consists of items from the Stanford 9 Achievement Test (SAT9) series as well as items developed by Delaware mathematics teachers. The DSTP is designed to measure progress towards the Delaware Content Standards, which were approved by the Delaware State Board of Education in 1995. "The mathematics section of the DSTP reflects that success in mathematics depends on a student's ability to grasp key concepts and solve realistic problems. Multiple choice questions, short answer questions, and extended response questions are used to assess students' conceptual knowledge, procedural knowledge, and knowledge of mathematical processes across core areas such as computation, measurement, algebra, and geometry." (http://www.doe.state.de.us/aab/DSTP_intro.html, p.3). A sample of released mathematics items as well as other information about the DSTP is available online at the Delaware Department of Education website. The reliability of the mathematics portion of the fifth grade DSTP in 2000 was .91 and the standard error of measurement was 11.7. The reliability of the mathematics portion of the eighth grade DSTP in 2003 was .92 and the standard error of measurement was 11.0.

Each student receives a scale score and a corresponding performance level. There are five possible performance levels. On the mathematics portion of the 8th grade DSTP students' scale scores could range from a minimum of 250 to a maximum of 750. On the mathematics portion of the 5th grade DSTP students' scale scores could range from a minimum of 175 to a maximum of 700. The relationship between the five performance levels at 5th grade and the corresponding scale scores are included in Table 2. Students in this study represent 5th grade children across all five performance levels. The distribution of students in this study is similar to Delaware children statewide (see Table 3). However, the percent of students in the highest performance level in mathematics are underrepresented in our sample most likely due to the exclusion of students enrolled in traditional 8th grade Algebra classes from this study.

Table 2

Performance Levels and Corresponding Scale Scores

Performance Levels	5th Grade Scale Score Categories
Well Below	423 and below
Below	424-448
Meets	449-502
Exceeds	503-524
Distinguished	525 and above

Table 3

Percent of Students by 5th Grade Math Performance Level

	Well Below	Below	Meets	Exceeds	Distinguished
Study Participants	18%	25%	52%	4%	1%
Delaware Students in 2000	(17%)	(21%)	(49%)	(8%)	(6%)

Quality of Mathematics Instruction

The framework used to determine the quality of mathematics instruction is the *Innovation Configuration* (IC) component of the Concern-Based Adoption Model (C-BAM) developed by Hall and Hord (1987). Hall and Hord discovered that when teachers begin to implement new innovations in their classrooms, highly effective implementation does not happen immediately. Typically, teachers move through a series of stages as they begin to work with and to better understand the innovation. An *Innovation Configuration* model is an instrument that depicts the continuum that teachers move through as they expand their use of the innovation. The instrument is based upon the actual patterns of implementation that have been observed in classrooms across the innovation. The process of developing the *Innovation Configuration*, which involved not only project staff and several Secondary Mathematics Specialists (SMS), but also the external evaluator, provides the conceptual framework for determining the quality of mathematics instruction for this study. The *Innovation Configuration* instrument developed for use in this study to gather information about the quality of mathematics instruction in individual middle school classrooms was adapted from the Local Systemic Change Classroom Observation Protocol from Horizon Research, Inc. in Chapel Hill, North Carolina.

Each of the 39 participating 8th grade classroom teachers received a global rating of high, moderate, or poor denoting the quality of their mathematics teaching. To compute this rating, each teacher first received a rating for each of the 17 traits (see Table 4) within the design and implementation, content, and classroom culture of the mathematics instruction. For each trait, classroom instruction was rated on a five point scale where "1" represented demonstration of high quality standards-based mathematics instruction and "5" represented demonstration of poor quality standards-based mathematics instruction. An example of the descriptions representing a "1", a "3", and a "5" from this instrument is listed in Table 5. The global rating was then determined by using the statistical procedure of cluster analysis to empirically form the three groups of teachers. The goal of the cluster analysis was to maximize homogeneity within groups and to maximize the heterogeneity between groups. This was accomplished using a k-means clustering technique in the statistical computer program, SPSS.

Each rater participated in the development of the *Innovation Configuration* model as well as attended numerous training sessions facilitated by the lead evaluator using the Annenberg video series, *Teaching Math: A Video Library 5-8*, to confirm the accuracy of the model in ascertaining the quality of standards-based mathematics instruction in each middle school classroom across Delaware. The development process was conducted over many sessions from March 2002 to November 2003 to develop and refine the instrument to ensure acceptable levels of validity and reliability.

Validity. The primary purpose of the *Innovation Configuration* instrument is to gather information on the quality of mathematics instruction provided in a middle school classroom during a single school year. To gather this information, the individual (or pair of individuals) with the most first-hand information about the quality of instruction in a particular classroom was selected to provide the data. For this study, eight raters were selected and trained to provide this information. Each rater was responsible for completing the *Innovation Configuration* instrument for about five to six teachers. In most cases, the sources of information included classroom observations, coaching sessions with the individual teacher, as well as interactions during professional development sessions. The 17 components of the instrument were selected from the classroom observation protocol developed by Horizon Research, Inc. for the Local Systemic Change through Teacher Enhancement (LSC) initiative funded by the National Science Foundation. This instrument is intended to reflect current standards in mathematics for exemplary practice as outlined in the National Council of Teachers of Mathematics' Principles and Standards for School Mathematics (2000).

Table 4

Components of the Innovation Configuration Instrument

Design and Implementation of the Mathematics Lesson
1. The teacher effectively engages students with important ideas related to the focus of the lesson. 2. The teacher provides adequate time and structure for investigation and exploration. 3. The teacher provides adequate time and structure for "wrap-up." 4. The teacher achieves a collaborative approach to learning. 5. The teachers' questioning strategies enhance the development of student conceptual understanding or sense-making. 6. The teacher is anchored in a clearly defined and communicated purpose. 7. The teacher is able to determine the students' level of understanding and adjust instruction accordingly.
Mathematics Content
8. The content is significant and worthwhile. 9. The content is challenging and developmentally appropriate. 10. The teacher provides content information that is accurate. 11. Elements of mathematical abstraction are included when appropriate to do so. 12. Appropriate connections are made to other disciplines and/or to real world contexts.
Classroom Culture
13. Active participation of all students is expected and valued. 14. There is a climate of respect for students' ideas, questions, and contributions. 15. The teacher's classroom management style/strategies enhance productivity. 16. The classroom climate encourages students to generate ideas, questions, conjectures, and/or propositions. 17. Intellectual rigor, constructive criticism, and/or the challenging of ideas are evident.

Table 5

Example of Teacher Behaviors Indicative of the Three Score Points on the Innovative Configuration Instrument

The teacher's questioning strategies enhance the development of student conceptual understanding or sense-making.	Score
✓ Teacher frequently asks students questions that require students to synthesize information, generate hypotheses, or form generalizations.	1
✓ Teacher asks students questions that are open-ended, but tone of voice or body language implies the teacher is looking for a specific response. Teacher curtails student responses.	3
✓ Teacher asks students questions that require students to respond with a one- or two-word answer. Teacher waits only a few seconds for a response.	5

Reliability. Both inter-rater reliability and internal consistency reliability for this instrument were calculated. Internal consistency reliability (coefficient alpha) was .91 during training. Coefficient alpha for the data used in this study was .97. These values indicate very good internal consistency for all traits on the instrument.

When individuals are trained on this instrument, we achieved an average inter-rater reliability of .77 with a range of .50 to 1.00. The inter-rater reliability rate for agreement is calculated to include values within one point of each other. Inter-rater reliabilities for acceptable agreement for each of the 17 elements are listed in Table 6. While a satisfactory level of reliability depends on how a measure is being used, in the early stages of a research study using instruments that have only a modest reliability, e.g., .70, is acceptable (Nunnally & Bernstein, 1994).

Table 6
Inter-Rater Reliability Statistics

Design and Implementation of the Mathematics Lesson	Inter-rater Agreement (± 1 pt.)
1. The teacher effectively engages students with important ideas related to the focus of the lesson.	70%
2. The teacher provides adequate time and structure for investigation and exploration.	70%
3. The teacher provides adequate time and structure for “wrap-up.”	70%
4. The teacher achieves a collaborative approach to learning.	80%
5. The teachers’ questioning strategies enhance the development of student conceptual understanding or sense-making.	89%
6. The teacher is anchored in a clearly defined and communicated purpose.	70%
7. The teacher is able to determine the students’ level of understanding and adjust instruction accordingly.	67%
Mathematics Content	
8. The content is significant and worthwhile.	50%
9. The content is challenging and developmentally appropriate.	70%
10. The teacher provides content information that is accurate.	77%
11. Elements of mathematical abstraction are included when appropriate to do so.	89%
12. Appropriate connections are made to other disciplines and/or to real world contexts.	90%
Classroom Culture	
13. Active participation of all students is expected and valued.	90%
14. There is a climate of respect for students’ ideas, questions, and contributions.	70%
15. The teacher’s classroom management style/strategies enhance productivity.	100%
16. The classroom climate encourages students to generate ideas, questions, conjectures, and/or propositions.	70%
17. Intellectual rigor, constructive criticism, and/or the challenging of ideas are evident.	90%

Limitations of this Research Design

Cumulative Effects. The design of this pilot study included measuring the quality of instruction students received in 8th grade. While measuring the quality of mathematics instruction students receive in the year in which they are assessed is valuable, the potential cumulative effects of the quality of instruction are missed. Research studies have shown the effects of the quality of instruction on student performance are additive and cumulative over grade levels (Wright, Horn, & Sanders, 1997). Therefore, while expensive, future studies will benefit if the quality of mathematics instruction is measured across multiple years for each student.

Instrument Development and Training. While the use of an independent criterion measure for the quality of instruction is critical to untangling the relationship between quality of instruction and student performance (Kupermintz, 2003), the creation of a valid, reliable instrument to measure the quality of instruction is a complex and time-consuming process. Therefore, creating an instrument with acceptable levels of inter-rater reliability requires the development of an instrument with clear and explicit scoring criteria. It also requires the use of raters who have expertise in mathematics and have received sufficient training in the appropriate use of the instrument (Herman, Aschbacher, & Winters, 1992). Although the design of this study incorporated the components necessary for high levels of inter-rater reliability, continued and additional training with knowledgeable raters using classrooms rather than video tapes would likely strengthen the inter-rater reliability of the *Innovation Configuration* instrument.

Sample size. The total number of students in this sample is sufficiently large for the analyses conducted. However, due to the small number of students in the sample that received a rating of "exceeds" or "distinguished" on the mathematics portion of the fifth grade DSTP, evidence of the impact of the quality of mathematics instruction on student performance for these students is limited. Therefore, claims about the impact of the quality of instruction on the performance of very strong math students should be made with caution.

Retrospective Data Collection. This study consists of a retrospective analysis of DSTP performance data from an existing population of eighth grade students and the quality of their educational experiences in mathematics. While the nature of classroom instruction is more difficult to document in hindsight, this study focused on gathering structured evidence rather than opinions regarding the quality of instruction by identifying specific teacher and student behaviors observed over the previous academic year. Future studies would benefit from supplementing this evidence by using the *Innovation Configuration* instrument to collect data during classroom observations, rather than retrospectively.

FINDINGS AND DISCUSSION

High Quality Mathematics Instruction

The mathematics instruction that took place in classrooms throughout the state of Delaware across 39 teachers involved in this study were clustered into one of three categories – high quality instruction, moderate quality of instruction, and poor quality of instruction. In Table 7, mean scores for each of the 17 traits is presented and disaggregated by quality of instruction. For each trait, an independent sample t-test was conducted to identify any significant differences among the means for the three groups regarding the quality of instruction. Across all 17 traits, there was a significant difference between at least two of the three groups. To identify where the differences were, a Scheffé post hoc test was conducted. Across nine of the traits, the difference among the three levels of the quality of instruction was significant. Among the remaining traits, nearly all indicated a significant difference between the high quality instruction classroom and less than high quality instruction classrooms. Therefore, many of these traits are important characteristics in differentiating among the quality of mathematics instruction.

Table 7

Mean Score of Seventeen Traits of Quality of Mathematics Instruction by Cluster

Trait		Quality of Instruction Cluster		
		High Quality	Moderate Quality	Poor Quality
Design and Implementation				
1.	Engages*	1.24	3.03	4.33
2.	Adequate*	1.79	3.53	4.67
3.	Wrap-up*	2.21	3.28	4.67
4.	Approach*	2.04	3.63	5.00
5.	Questioning**	2.03	3.75	4.50
6.	Anchored*	1.00	2.72	4.42
7.	Adjust**	1.91	3.70	4.42
Mathematics Content				
8.	Content*	1.24	2.47	4.20
9.	Appropriate*	1.62	3.47	4.30
10.	Accurate***	1.00	1.75	4.10
11.	Elements*	2.54	3.84	4.80
12.	Connections*	1.50	3.00	4.30
Classroom Culture				
13.	Active**	1.62	3.33	3.34
14.	Climate**	1.90	3.56	4.00
15.	Style**	1.33	2.93	3.67
16.	Ideas**	1.93	3.13	3.75
17.	Rigor**	1.94	3.66	4.67

* Significant differences ($p<.05$) among all three means.

** Significant difference ($p<.05$) between mean of high quality instruction and mean of poor quality instruction.

*** Significant difference ($p<.05$) betweens mean of moderate quality instruction and mean of poor quality instruction.

Distribution of Students

Most of the students in this study were receiving high (43%) or moderate (46%) quality mathematics instruction in 8th grade as measured by the *Innovation Configuration* instrument. However, students were not equally likely to receive high quality instruction in mathematics in 8th grade. Figure 1 and Table 8 indicates that students who received poor quality mathematics instruction in 8th grade, on average, were somewhat weaker mathematics students in 5th grade.

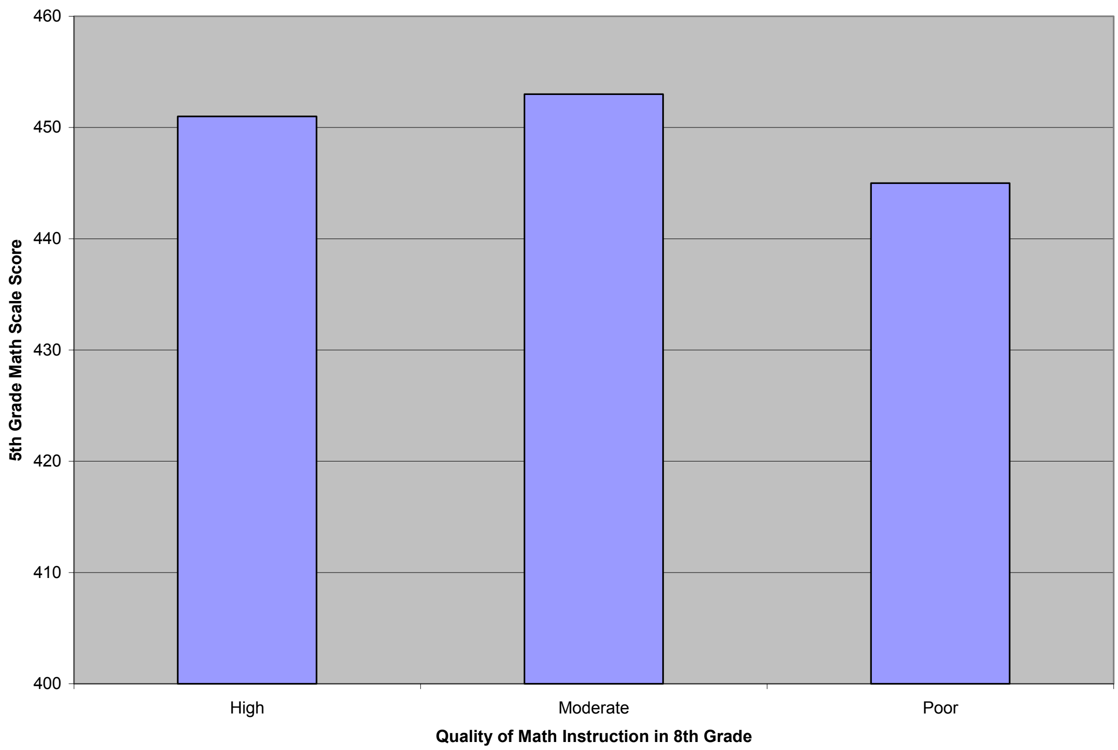


Figure 1. Mean scale score of students by quality of instruction.

Table 8

Mean Scale Score of Students by Quality of Instruction

	Quality of Instruction in 8th Grade		
5th Grade Math DSTP	High	Moderate	Poor
Mean Scale Score	451.18	452.81	445.47
Standard Deviation	31.17	31.33	33.33
Sample size	900	958	244

While all three categories of the quality of instruction (high, moderate, and poor) contained strong as well as weak math students, the distributions were slightly different (see Figure 2). For example, students who scored “well below” on the mathematics portion of the 5th grade DSTP (n=383) are represented in greater proportions in classrooms where poor quality of instruction occurred (see Table 9).

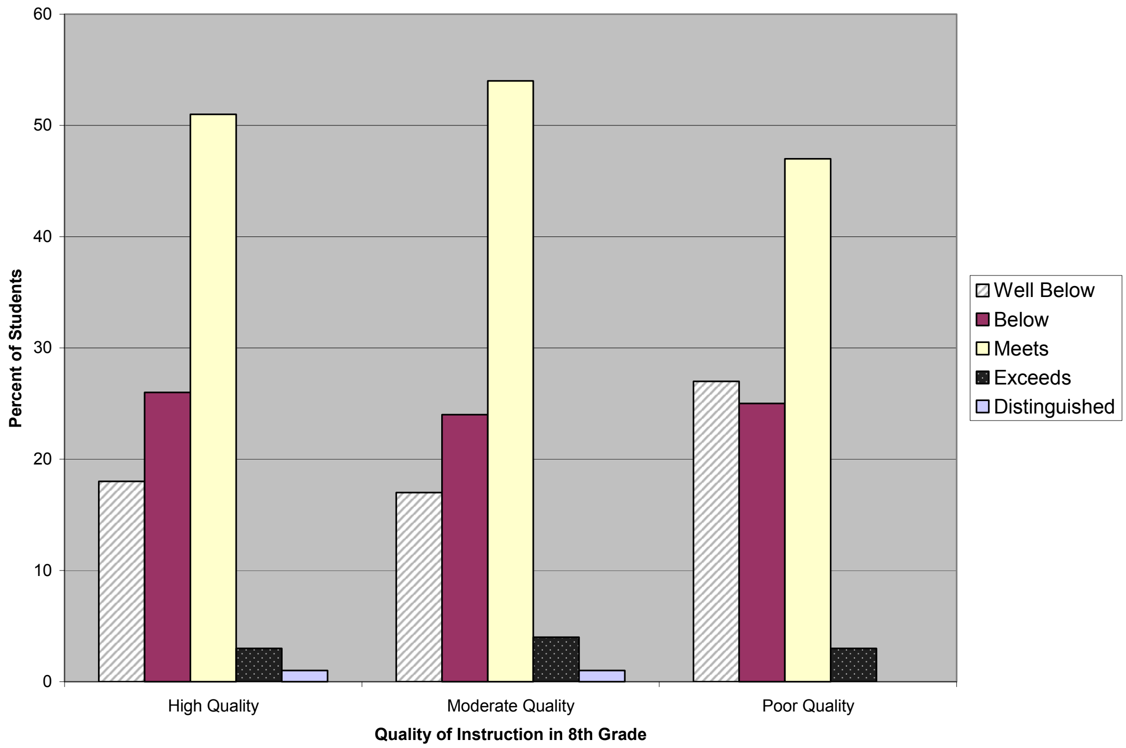


Figure 2. Percent of students in each performance level by quality of instruction.

Table 9

Percent of Students in Each Performance Level by Quality of Instruction

	Quality of Mathematics Instruction in 8th Grade		
Performance Level on 5th Grade DSTP	High Quality	Moderate Quality	Poor Quality
Well Below	18%	17%	27%
Below	26%	24%	25%
Meets	51%	54%	47%
Exceeds	3%	4%	3%
Distinguished	1%	1%	0%

Twenty-seven percent (27%) of the students who received poor quality of instruction in 8th grade, received a performance level of "well below" on the mathematics portion of the 5th grade DSTP. However, only 18% of the students who received high quality instruction in 8th grade received a performance level of "well below" on the mathematics portion of the 5th grade DSTP. Therefore, students who received a performance level of "well below" the standard on the 5th grade DSTP were 1.47 times more likely than other 8th grade students to be placed in a classroom where they received poor mathematics instruction than to be placed in a classroom where they received moderate or high quality mathematics instruction in 8th grade. In addition, none of the students who earned a performance level of "distinguished" on the math portion of 5th grade DSTP (n=22) received poor quality mathematics instruction in 8th grade.

Quality of Instruction and Student Achievement

Because students were not equally likely to receive high quality mathematics instruction in 8th grade, a two-way ANOVA rather than a one-way ANOVA was conducted to factor in previous mathematics achievement. The 3 (quality of instruction) x 5 (math performance levels in 5th grade) between groups ANOVA revealed a main effect for previous math achievement, $F(4, 2088) = 230.378$, $p < .001$, and a main effect for quality of instruction, $F(2, 2088) = 3.192$, $p < .05$. There was no significant interaction between quality of instruction and previous math achievement, $F(7, 2088) = .634$, $p = .73$. This indicates that not only does prior math achievement explain differences in future performance in mathematics, but the quality of instruction students received also explains some of the differences. Together, these two variables, prior math achievement and quality of instruction, explained 39% of variance among scale scores on the mathematics portion of the 8th grade DSTP.

Quality of Instruction. The means for each of three groups shown in Figure 3 and Table 10 indicate the positive relationship between the quality of instruction and math achievement in eighth grade. That is, students who received high quality of instruction in 8th grade performed significantly better on the mathematics portion of the DSTP than students who received poor quality of instruction in the 8th grade.

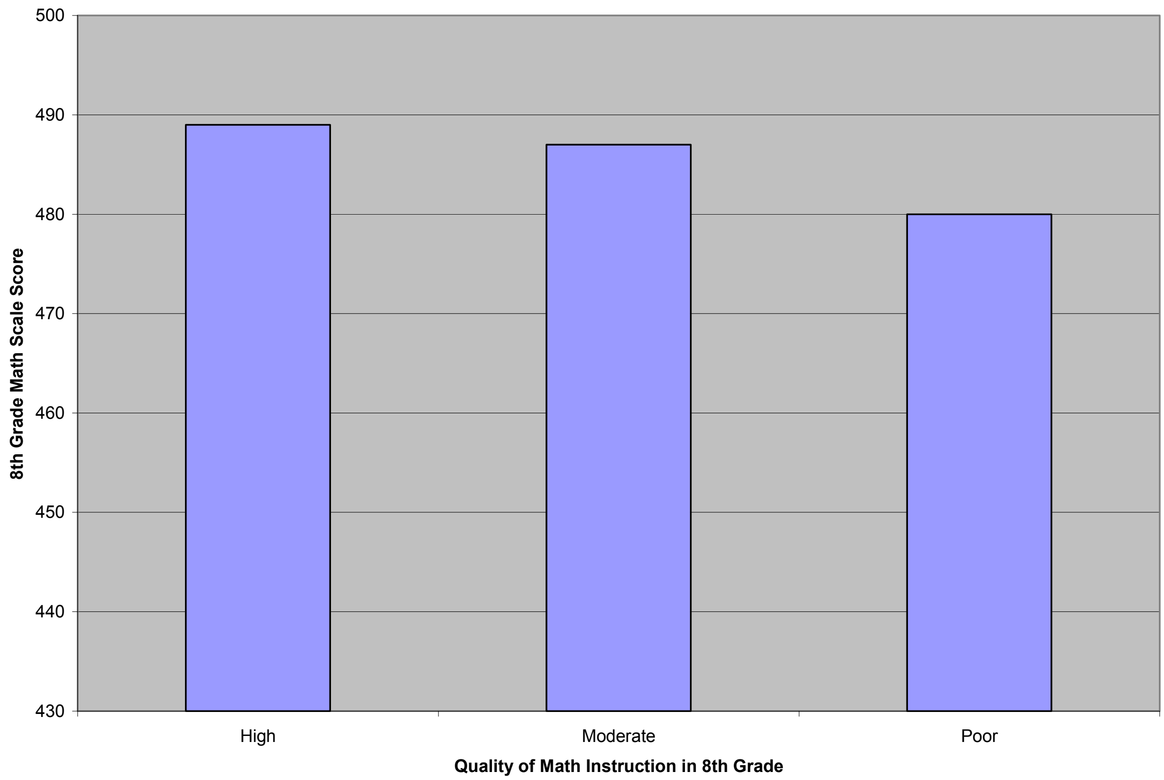


Figure 3. Mean scale scores by quality of instruction.

Table 10

Mean Scale Scores by Quality of Instruction

	Quality of Instruction in 8th Grade		
	High	Moderate	Poor
Mean Math Scale Score (8th grade)	489.42	486.67	480.22
Standard Deviation	27.13	26.71	24.01
Sample Size	900	958	244

Previous Achievement. The means for each of five groups shown in Figure 4 and Table 11 indicate the positive relationship between previous achievement and math achievement in eighth grade. That is, students who performed well in 5th grade performed significantly better on the mathematics portion of the 8th grade DSTP than students who performed poorly in 5th grade.

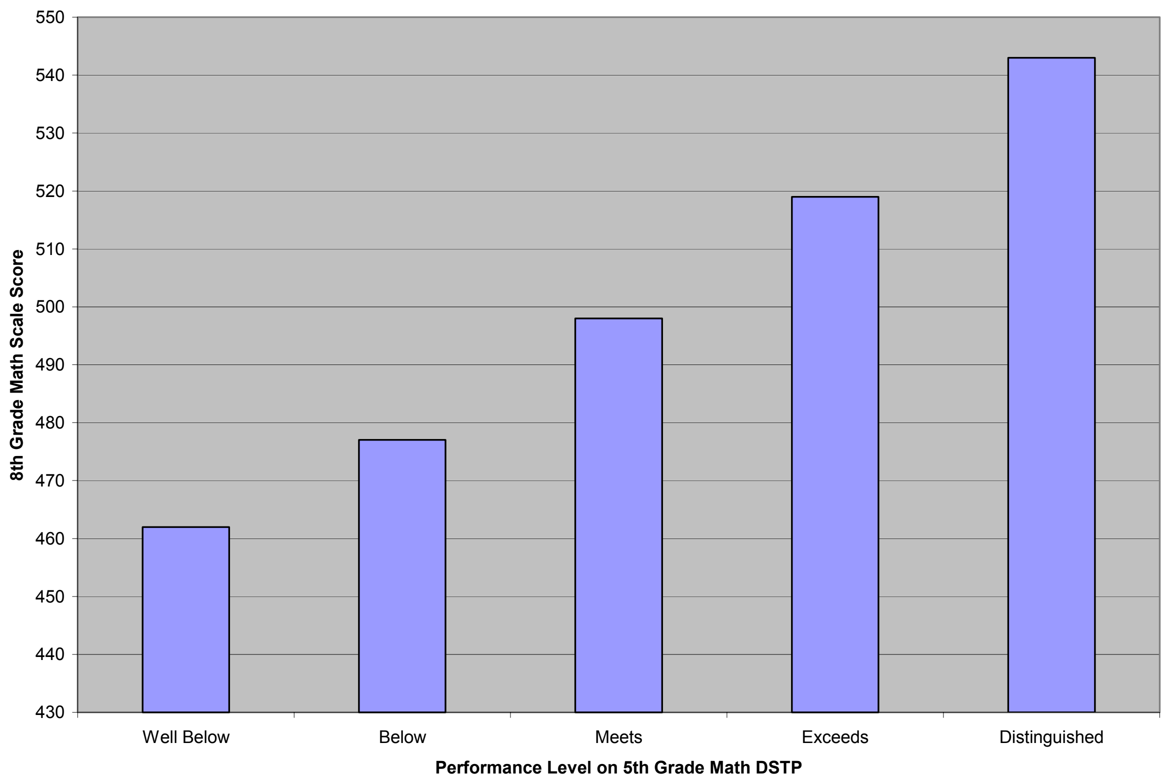


Figure 4. Mean scale score of students with various performance levels.

Table 11

Mean Scale Scores of 8th Grade Students with Various Performance Levels

	Performance Level in 5th Grade				
	Well Below	Below	Meets	Exceeds	Distinguished
Mean Math Scale Score	461.93	476.57	497.62	519.05	543.45
Standard Deviation	19.22	19.19	22.01	21.61	36.69
Sample Size	383	527	1093	77	22

For the purposes of this study, we are mostly interested in differences in student performance based on the quality of instruction students received. Figure 5 shows the mean scale scores on the mathematics portion of the 8th grade DSTP for each group of students. This graph indicates that students who received high quality instruction in 8th grade and received a performance level of "below," "meets," or "exceeds" in 5th grade performed better in 8th grade than students who received poor quality instruction in 8th grade. On average, the students who received high quality instruction earned a mathematics scale score 8 to10 points higher than those who received poor quality instruction.

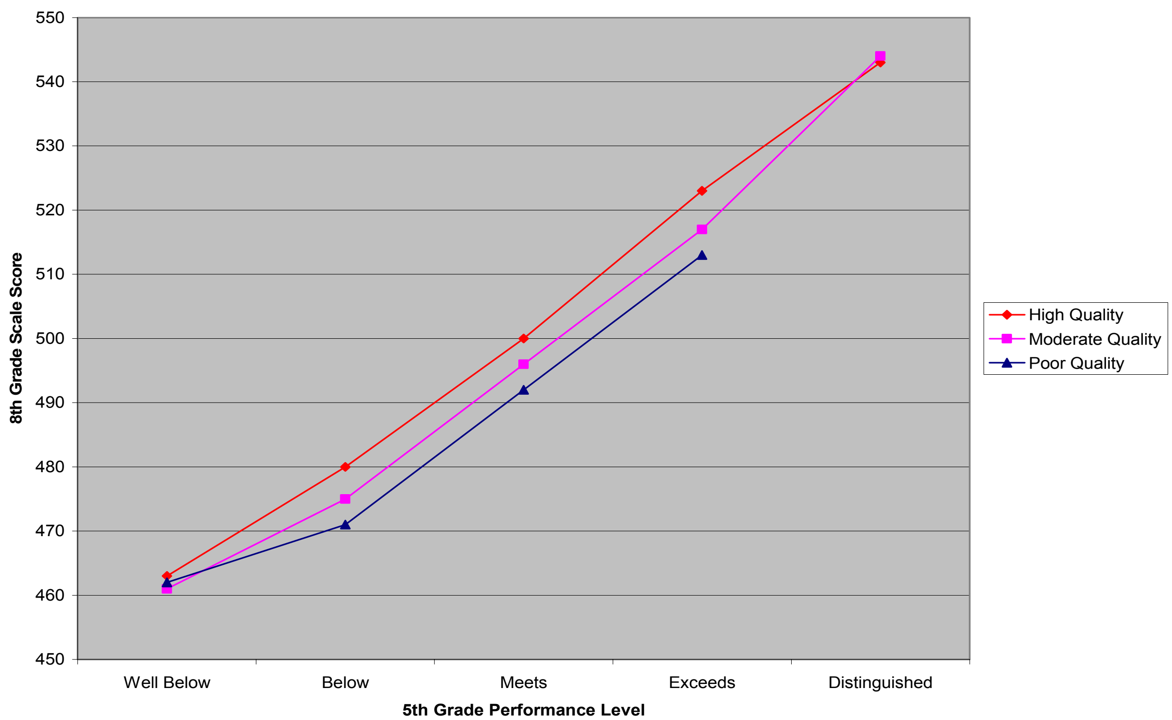


Figure 5. Mean mathematics scale score by quality of instruction in 8th grade and performance level

Table 12

Mean Scale Scores on the 8th grade DSTP by Quality of Instruction and Prior Achievement

Performance Level on 5th Grade Math DSTP	Sample Size	Quality of Instruction in 8th Grade		
		High Quality	Moderate Quality	Poor Quality
Well Below	383	462.97	460.96	461.62
Below	527	479.73	474.88	471.21
Meets	1093	500.23	496.48	492.24
Exceeds	77	523.30	516.95	513.38
Distinguished	22	542.55	544.36	--

To identify any *significant* differences, confidence intervals were calculated for each of the 14 means listed in Table 12. If the confidence intervals overlap, the difference between the groups is not significant. However, if the confidence intervals do not overlap, the difference between the groups is significant. The 14 confidence intervals are listed in Table 13. Comparing the mean scale scores on the 8th grade DSTP of each of the 14 groups of students indicate that a few comparisons were statistically significant.

Table 13

Confidence Intervals on Scaled Scores on the 8th Grade DSTP by Quality of Instruction and Performance Level

Performance Level on 5th Grade DSTP	Sample Size	Quality of Instruction in 8th Grade		
		Poor Quality	Moderate Quality	High Quality
Well Below	383	454.2 - 464.2	457.7 - 464.2	459.8 - 466.2
Below	527	466.4 - 476.6	472.2 - 477.6	477.0 - 482.4
Meets	1093	488.4 - 496.1	494.7 - 498.3	498.3 - 502.1
Exceeds	77	498.9 - 527.9	510.4 - 523.5	515.8 - 530.8
Distinguished	22	--	532.0 - 556.8	530.2 - 554.9

Students who scored in the middle. The greatest impact occurred for students who scored in one of middle three categories (below, meets, or exceeds) in 5th grade. For example, students who scored "below the standard" or "meets the standard" on the 5th grade DSTP and received high quality mathematics instruction in 8th grade performed significantly better on the 8th grade DSTP than those who received poor quality mathematics instruction in 8th grade (see Figure 5). There does not appear to be a relationship between the quality of instruction and student achievement for students whose performance exceeds the standard in 5th grade (see Table 13). However, due to the small sample size for this category of students, a statistically significant difference is much more difficult to detect. Thus, a larger sample of students for this group would be necessary to determine if there is a relationship between the quality of instruction and performance in 8th grade for students who "exceeded the standard" in 5th grade.

Students who scored at the extremes. There was no significant difference between performance of students who received high quality instruction and those who received poor quality instruction for students who scored "well below the standard" on the 5th grade DSTP. That is, there does not appear to be a relationship between quality of instruction and math achievement in 8th grade for students whose performance is well below the standard in 5th grade. One hypothesis is that this lack of difference could be attributed to a situation of "too little, too late."

Furthermore, there does not appear to be a relationship between quality of instruction and math achievement in 8th grade for students whose performance is "distinguished" in 5th grade. For these students the difference in the average means is only one to two scale score points. In this case, one hypothesis for this lack of difference could be attributed to a situation in which students perform well regardless of the quality of their educational experience due to other out-of-school factors. In addition, it is worth noting that none of the students who were classified as distinguished on the 5th grade mathematics portion of the DSTP received poor quality of instruction in 8th grade. Given the small sample size for this group of students, it is not possible in this study to determine if this happened by chance or by design.

CONCLUSIONS

While previous mathematics achievement is a significant predictor of future performance, this study indicates that for many students there is a significant relationship between the quality of instruction a student receives and his/her mathematics achievement. We found that students performing at “below” or “meets the standard” on the 5th grade state assessment who received high quality mathematics instruction performed significantly better than their peers who received poor quality mathematics instruction. Therefore, for most 8th grade students, the quality of mathematics instruction they receive in 8th grade seems to make a difference.

However, this study also indicates that all students do not have an equal chance of being assigned to an 8th grade classroom where they will experience high quality instruction. A student who performs very poorly in mathematics in 5th grade is more likely to be assigned to a classroom in which he/she will receive poor quality mathematics instruction. A student who performs very well in mathematics in 5th grade is less likely to be assigned to a classroom in which he/she will receive poor quality mathematics instruction. While the reasons for this imbalance cannot be explained by this study, the mere existence of it confounds the problem of poor performance.

References

Aronson, S. R., Mutchler, S. E., & Pan, D. T. (1998). *Theories of Change: Making programs accountable and making sense of program accountability.* Austin, TX: Southwest Educational Development Laboratory.

Assessment and Analysis Group (2001). *Delaware Student Testing Program Technical Report 1998-2000.* Dover, DE: Delaware Department of Education, Assessment and Accountability Branch.

Assessment and Analysis Group (2003). *Delaware Student Testing Program Online Summary Reports 2003.* Dover, DE: Delaware Department of Education, Assessment and Accountability Branch.

Hall, G. E., & Hord, S. M. (1987). *Change in Schools.* Albany, NY: State University of New York Press.

Herman, J. L, Aschbacher, P. R. & Winters, L. (1992). *A Practical Guide to Alternative Assessment.* Alexandria, VA: Association for Supervision and Curriculum Development.

Keith, P. B. & Lichtman, M. (1992, April). *Testing the influence of parental involvement on Mexican-American Eighth Grade Students' academic achievement: A structural equations analysis.* Paper presented at the meeting of the American Educational Research Association , San Francisco, CA.

Nunnally, J. C. & Bernstein, I. H. (1994). *Psychometric Theory.* New York: McGraw-Hill, Inc.

Sanders, W. L. (1998). *Value Added Assessment.* The School Administrator, December, pp. 24-27.

Sanders, W. L. & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement.* Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center.

Vollmer, F. (1986). The relationship between expectancy and academic achievement—How can it be explained? *British Journal of Educational Psychology, 56,* 64-74.

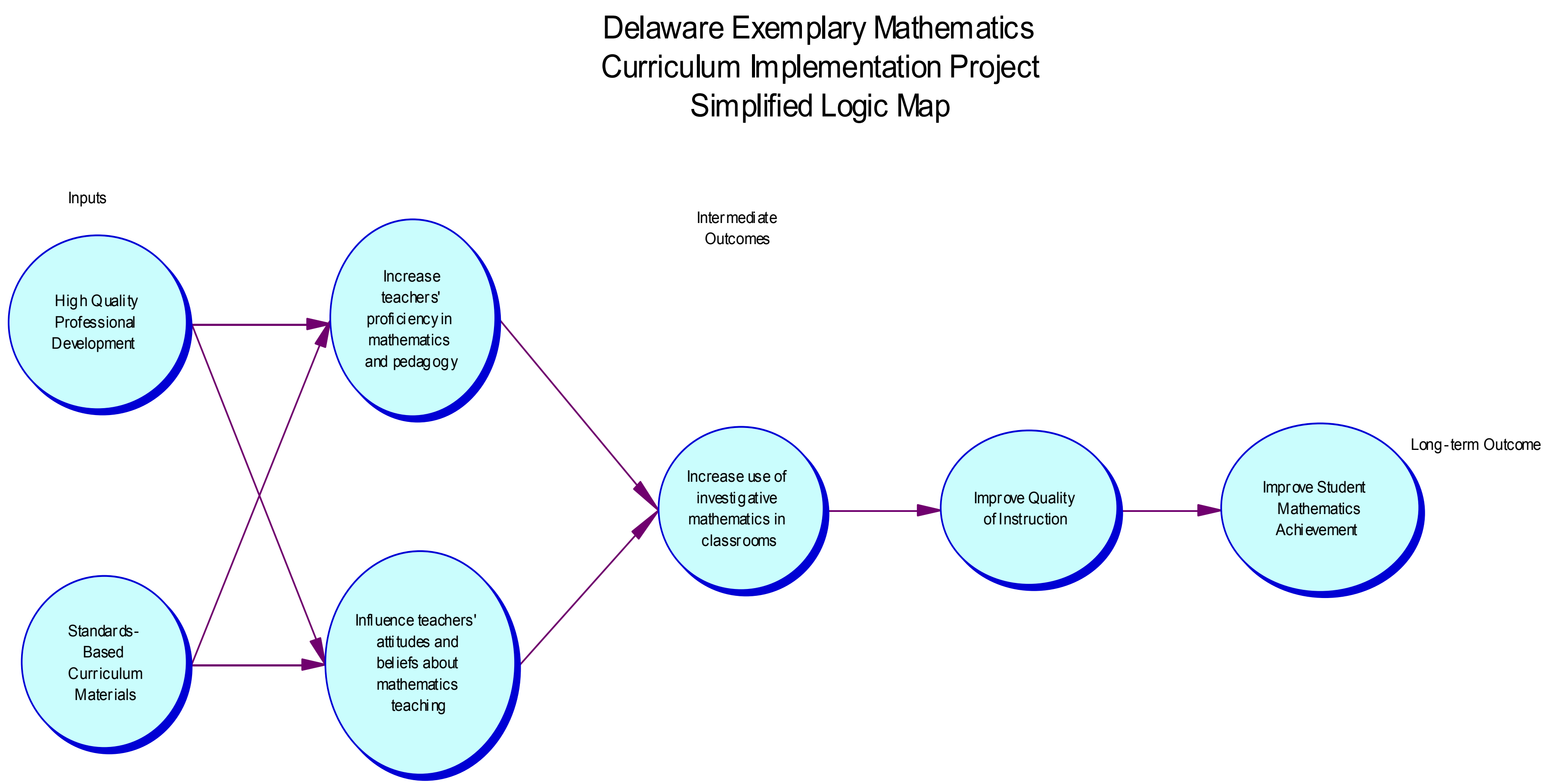
Weiss, C. H. (1995). Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families. In J. P. Connell, A. C. Kubisch, L.B. Schorr, and C. H. Weiss (eds.), *New approaches to evaluation community initiatives: Concepts, methods, and contexts.* Report by the Roundtable of Comprehensive Community Initiatives for Children and Families. Washington, DC: The Aspen Institute.

Wright, S. P., Horn, S. P., & Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: Implication for teacher evaluation. *Journal of Personnel Evaluation in Education, 11,* 57-67.

Young, D. J., Reynolds, A. J., & Walberg, H.J. (1996). Science achievement and educational productivity: A hierarchical linear model. *Journal of Educational Research, 89,* 272-78.

Appendix A

Theory-Based Logic Map


Delaware Exemplary Mathematics
Curriculum Implementation Project
Simplified Logic Map
Inputs
Intermediate Outcomes
Long-term Outcome
High Quality Professional Development
Standards-Based Curriculum Materials
Increase teachers' proficiency in mathematics and pedagogy
Influence teachers' attitudes and beliefs about mathematics teaching
Increase use of investigative mathematics in classrooms
Improve Quality of Instruction
Improve Student Mathematics Achievement

Acknowledgements

Many thanks to the following people for their time and effort:

Jennifer Burlew

Margaret Byrem

Sally Caldwell

Nicole Durkin

Kathleen Hollowell

Jane Mahoney

Jon Manon

Valerie Maxwell

Gina Moody

Jamila Riser

Mary Lynn Vincent

I would like to extend my thanks to the Secondary Math Specialists and the staff at the Mathematics and Science Education Resource Center for sharing their thoughts, opinions, time and talents. I would also like to extend my thanks to Sally Caldwell at the Delaware Department of Education for her assistance with the student achievement data for this study. Without their valuable input and assistance, this report would not have been possible.

~ Thank you